FILE

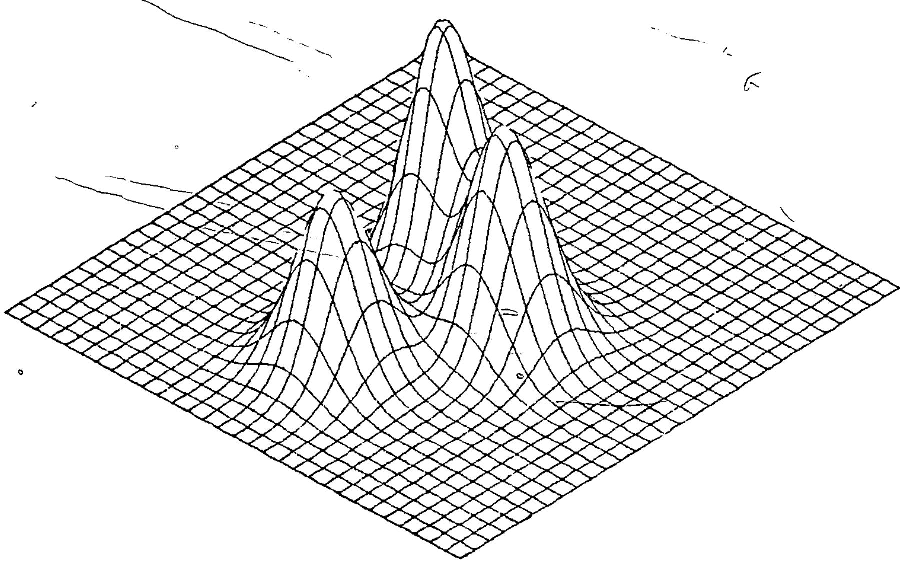
SECRET

25X1

ANALYTIC METHODS

Directorate of Intelligence
Central Intelligence Agency

DATE 12/5/86 FILE 3
DOC NO ASG-M 86-20012
OIR 3 5-7
P & PD 1 4
DESTROYED 2 copy



SECRET

This document is a research paper by the Analytic Methods Branch of the Directorate of Intelligence of the Central Intelligence Agency. As such, the views and conclusions contained herein are those of the research staff and should not be interpreted as necessarily representing the official position, either expressed or implied, of the Central Intelligence Agency.

SECRET

25X1

Directorate of Intelligence
Central Intelligence Agency

May 1986

SUMMA AGRICULTURAE: STATISTICAL ISSUES IN AGRICULTURAL CROP ASSESSMENTS IN DEVELOPING COUNTRIES

Analytic Support Paper

Prepared for

Office of Global Issues

Prepared by

25X1

ASGM 86-20012
Report No. 275
Copy No. 3

25X1

Comments are welcome and may be addressed to the
Chief, Analytic Methods Branch, ASG,

25X1

SECRET

25X1

SUMMA AGRICULTURAE: STATISTICAL ISSUES IN AGRICULTURAL CROP ASSESSMENT IN DEVELOPING COUNTRIES

ABSTRACT

Assessing the cultivation of agricultural crops, from the grain crop in Afghanistan to the effect of the drought in Ethiopia, has been a major intelligence problem for many years. An interdisciplinary approach to crop estimation, using statistical analysis of imagery-derived data, provides a quantitative evaluation of these issues. The statistical procedure involves designing a sample, estimating the number of fields, average field size, and if possible, combining these with yield estimates to calculate a final production estimate. A major area for future study is to develop improved sampling techniques for acquiring the imagery. Finally, better assessments of yield are required for estimating total production.

Analytic Methods Branch
Analytic Support Group
ASGM 86-20012
Report No. 275
May 1986

25X1

"Data! data! data!" he cried impatiently.
"I can't make bricks without clay."

- Sherlock Holmes

25X1

TABLE OF CONTENTS

LIST OF TABLES

25X1

LIST OF FIGURES

25X1

INTRODUCTION

The recent drought and food shortages in Africa dramatize the importance of agriculture to the general well-being and political stability of the region. These problems, while magnified in places like Ethiopia, are common to many developing nations. The burdens of a growing population, coupled with outdated farming methods, make it difficult to provide the basic necessities in parts of the developing world. Timely and reliable assessments of agricultural output are, therefore, essential to US policymakers who must address US aid to these countries and related issues.

This paper presents an overview of the statistical methodology used in agricultural crop assessments in developing countries, underscoring the data required and the techniques used to obtain the final crop estimate. Due to the wide variety of agricultural estimation problems, ranging from grains in Afghanistan and Ethiopia to citrus orchards in the Caribbean, no single method can be applied in each situation. However, similar components exist in most estimation problems and many of the statistical issues addressed here are common to all problems. Throughout the discussion, assumptions inherent

Sanitized Copy Approved for Release 2011/02/22 : CIA-RDP86T01017R000707650001-5

[redacted]

25X1

in the methodology are clarified and recommendations for improvement, where possible, are offered.

The statistical methods that are applicable to crops grown in developing countries differ from the estimation techniques used for Soviet grain (reference 11). In studying Soviet grain production, the area under cultivation is known and the major effort centers on determining the yields. The availability of LANDSAT imagery and published historical records supplement data collected by national reconnaissance assets. In the developing world, by comparison, even estimates of the hectareage sown in a particular crop are difficult to derive. The fields are usually too small to observe on LANDSAT, making higher resolution imagery the only reliable source of data. Published data are either not available or of questionable quality. In short, fundamental estimation problems must be resolved before crop assessments in the developing world will reach the level of refinement currently found in Soviet grain assessments.

While the methodology discussed here bears some resemblance to the techniques used to assess certain narcotics crops, important differences are worth noting.

25X1

Since narcotics cultivation is either illegal or highly regulated, most growers strive to avoid detection. Narcotics fields are usually in more isolated or remote areas. Such is not the case with food crops. The isolation of some narcotics crops, such as coca fields in Colombia, simplifies the identification of these fields on imagery. By comparison, several different food crops may be cultivated in adjacent fields and difficulties arise in correctly classifying the fields. Furthermore, there is more basic research available on cultivation and yields for food crops, as compared to drug crops. This knowledge should permit more precise information on yields, the crop calendar, and farming techniques. Finally, because narcotics crops are illegal, growers do not usually welcome outside visitors. Although it is difficult to visit some farming regions, food crops should be more accessable, at least in some countries.

Agricultural crop assessment includes its own set of problems, which any statistical methodology must address:

- Imagery used for agricultural assessment is often obtained as a result of a priority collection requirement against nearby military installations, thus sampling biases are difficult to assess.

- Sample sizes are often quite small. Therefore estimates obtained from these samples embody a wide range of uncertainty.
- Ground truth in countries like Afghanistan and Ethiopia is not available.

The methodology described in this report may be viewed as a "generic" estimation procedure. The next two sections discuss the estimation method and related issues of imagery acquisition. In application, these methods are tailored to the specific country and crop under consideration. Variations in the estimation methods may arise from the type of imagery, availability of ground truth, differences in cultivation techniques or other factors.

ESTIMATING CROP PRODUCTION

The essential idea underlying a crop estimate is that total production equals the area cultivated times the average yield, where the area cultivated is the total number of fields times the average field size. When yield figures are available for different categories of fields, a more accurate estimate can be obtained by multiplying each area by its average yield and summing the respective parts. Thus, to estimate production, the number of fields, the average field size, and the average yield must be estimated. Were each known precisely, production could be calculated precisely.. In reality, however, each figure must be estimated and some error arises in the process (figure 1). An examination of the steps in the estimation process will highlight the main problems and reveal the sources of possible uncertainty.

Overhead imagery or aerial reconnaissance provides the data for an estimate of the number of fields. If the entire potential growing area is imaged, then a full census is possible. Due to time and cost considerations, this may not be desirable. Usually, the available imagery constitutes a sample of the potential growing area, and the number of fields observed on the imagery is "scaled up" to account for the area not imaged (reference 2).

This scale-up procedure assumes that the potential growing area is known and that the imagery is a representative sample of the entire region. The potential growing area determines how much to inflate the number of fields counted on the imagery, in order to estimate the total number of fields. If the figure used for the potential growing area is too small, the total number of

25X1

FIGURE 1. PROCEDURE FOR ESTIMATING TOTAL PRODUCTION

PROCEDURE	SOURCE OF INFORMATION
Design the sample.	Statistician.
Count number of fields on imagery.	Imagery.
Scale up number of fields from imagery to total growing area, giving total number of fields.	Total growing area is based on historical information.
Adjust for cloud cover.	Imagery.
Adjust for detection probability.	Ground truth or calibration experiment.
Multiply total number of fields by average field size to obtain total hectares cultivated.	Average field size is calculated by mensuration of the imagery.
Multiply total hectares by yield per hectare to obtain the production estimate.	

25X1

25X1

fields will be underestimated. If the figure is too large, the total number of fields will be overestimated. Imagery that is not a representative sample of the growing region overall, will contain a density of fields that differs from the density of fields in the area not imaged, causing a bias in the scaled-up estimate for the total number of fields. This bias can be avoided by recognizing factors which influence growing potential and incorporating these factors into the sample design.

To illustrate the estimation procedure, suppose 100 square kilometers were imaged and 500 fields observed on the imagery (table 1, stratum[1] of intense cultivation). Often the area imaged is larger than the ground area observed on imagery because cloud cover obscures some of the region. The calculation of the number of fields per square kilometer must compensate for this problem. In this example, if there were no cloud cover, then the density would be five fields per square kilometer, as

[1]Sometimes the uncertainty in the final estimate can be reduced by separating the region into two regions or "strata" which are more homogeneous. Each stratum is analyzed separately and the results combined to obtain the final estimate. A discussion of this technique appears in the next section under "Designing the Sample".

25X1

TABLE 1. EXAMPLE OF CALCULATIONS FOR ESTIMATING CULTIVATION AND PRODUCTION

QUANTITY	STRATUM		TOTAL
	INTENSE CULTIVATION	LESS INTENSE CULTIVATION	
Potential Growing Area (km^2)	1,000	5,000	6,000
Area Imaged (km^2)	100	100	- - -
Fields Observed	500	200	- - -
Cloud Cover (percent)	20	30	- - -
Estimated Density of Fields (fields/km^2)	6.25	2.86	- - -
Scaled-up Estimate of No. of fields	6,250	14,300	- - -
Detection Probability (percent)	75	75	- - -
Corrected Estimate of No. of Fields	8,333	19,067	27,400
Average Field Size (ha)	0.5	0.3	
Total Area Cultivated (ha)	4,167	5,720	9,887
Yield (quintals/ha)	20	20	
Total Production (mt)	8,334	11,440	19,774

shown earlier. Suppose instead, that 20 percent of the film were cloud covered. Then only 80 square kilometers of the ground appeared on imagery and the density of fields is 500/80 = 6.25 fields per square kilometer. Since the potential growing area is 1,000 square kilometers, the estimated number of fields, correcting for cloud cover, is 6,250 (=1,000 x 6.25). This calculation is called "scaling up" because the number of fields observed on imagery represents only a fraction of the number of fields in the total growing area.

Counting the number of fields on imagery is a complex process which requires the development of suitable "keys" for identifying the crop under consideration. Typically, these keys are derived from a comparison of ground-truth photography with the imagery derived from overhead collection. Because of the variations in farming practices, the keys developed for one region are not generally transferable to another. As the imagery analyst (IA) gains experience in a region and a variety of keys are developed from various crop conditions, the quality of the imagery-based assessment will improve.

Suppose the crop under consideration is wheat. If all of the wheat fields were correctly identified on

[redacted]

25X1

imagery and no non-wheat fields were misidentified as wheat fields, then the value obtained above would be the final estimate of the number of wheat fields. Often, though, errors in detection and classification of fields arise from problems with the quality of imagery or uncertainty over the signature of the crop. The judicious use of ground truth makes it possible to correct for these difficulties (reference 2).

Comparison of ground truth photography to overhead imagery permits estimation of the "detection probability", which is the probability of correctly identifying a field on imagery. Failure to adjust for the detection probability will cause total production to be underestimated. In some instances, ground truth is not available and an alternative statistical tool, known as capture-recapture[2] modeling, is used to estimate the detection probability (references 1,3).

---

[2]Capture-recapture is a technique that was developed for estimating mobile animal populations, such as the number of fish in a lake. The Analytic Methods Branch, and its predecessor in ORD adapted this technique for use with several intelligence problems, including opium in Mexico, civil unrest in the Soviet Union, and Soviet arms transfers to Third World countries.

25X1

The misclassification probabilities are the probability of classifying a wheat field as a non-wheat field and visa-versa. Again, the comparison of ground truth to overhead imagery permits adjustments for errors in classification. The details of this mathematically complex step vary with the type of crop and the quality of the imagery.

Suppose, to continue the hypothetical example (table 1) that the IA correctly identifies 15 fields on several frames of imagery, but the ground-truth photography demonstrates that 20 fields were actually present. The detection probability, the chance of correctly identifying a field on imagery given that it is truly there, is 15/20 or 75 percent. The estimate of 6,250 fields calculated above should now be corrected for the 25 percent that were missed, giving a final estimate of 6,250/0.75 = 8,333 fields.

If ground truth were not available, the capture-recapture approach could be used to estimate the detection probability as follows. Suppose one IA examines several frames of imagery and identifies 15 fields. These fields are then marked on an overlay. A second IA then examines the same frames. Suppose he identifies 18 fields without knowing those fields identified by the first IA. On comparison, 13 fields were discovered by both IAs. Then the total number of fields is estimated to be (15 x 18)/13 = 20.76 and the detection probabilities are 15/20.76 = 72 percent and 18/20.76 = 87 percent for the two IAs, respectively (table 2).

The final estimate of the number of fields has a range of uncertainty associated with it due to the error in the estimation process. Generally statisticians consider the error to be in two parts: the bias, which arises from non-representative sampling, and the random variation which is due to only having a sample of the data. Biases are particularly serious, since it is generally impossible to assess their size. Careful planning of the data collection is essential in minimizing the bias in the final estimate. The size of the random variation can be estimated by the appropriate variance. Furthermore, this type of error can be minimized by sensible allocation of data collection resources. Methods for minimizing both types of error are addressed in the next section.

In addition to the number of fields, an estimate of average field size is necessary for assessing agricultural

25X1

TABLE 2. CAPTURE-RECAPTURE: THE NUMBER OF AGRICULTURAL FIELDS DETECTED ON IMAGERY BY TWO IMAGERY ANALYSTS (IA)

		FIRST IA		
		DETECTED	NOT DETECTED	TOTALS
SECOND IA	DETECTED	13	5	18
	NOT DETECTED	2	1	3
	TOTAL	15	6	21

SUMMARIZING THIS INFORMATION:

- The capture-recapture technique indicates there are roughly 21 fields on the imagery.
- The first IA identified 15 fields, giving a detection probability of 72 percent. Similarly, the second IA has a detection probability of 87 percent.

 25X1

crop production. Typically, the average field size is obtained by mensurating a sample of fields and averaging their values. Provided the fields mensurated are a random sample of all fields, the accuracy of this estimate depends on the variation among fields and the number of fields measured. The error in this step can be made as small as desired by taking a large enough random sample. The variance of the average field size is incorporated into the final variance estimate for the area cultivated. A detailed discussion of the procedure for calculating these variances is discussed in appendix A. To continue with the example (table 1), suppose the average field size is 0.5 hectares. Then the total area cultivated is the average field size times the estimated number of fields, namely 0.5 x 8,333 = 4,167 hectares.

[redacted] 25X1

[redacted] Total production is estimated as the product of the yield and the area cultivated. If, for example, the yield is 20 quintals per hectare then total production would be 20 x 4,167 = 83,340 quintals or 8,334 metric tons. Since only the number of fields and the average field size were derived through statistical analysis, only the uncertainty associated with the area cultivated can be 25X1 calculated. These calculations assume that the misclassification probabilities are zero.

It may be possible, in some cases, to obtain better estimates of yield through the development of a crop model or by some specialized method. For grain crops in the Soviet Union, for example, yields can be estimated from "straw dumps" (references 12, 13). An analogous procedure, based on shocks of straw observed on imagery, could be applied in some developing countries. The underlying concept is that grain yield is closely related to the total biomass. The volume of straw residue remaining from the harvest is a good indicator of total biomass and, therefore, of yields. The precise relationship between the number of straw shocks per hectare and the grain yield depends on the type of grain and the regional farming practices. Using either historical data, when available, or surrogate data from a similar agricultural area, the precise grain-to-straw relation can be calibrated.

When the growing region under study has been stratified, as in this example, the estimation procedure described above is performed on each stratum separately and the final estimate is the sum of the estimates for the two strata (table 1). To obtain the variance for the final estimate, calculate the variance for each stratum and add the two. Note that the confidence interval for the final estimate is not the sum of the confidence intervals for the two strata. This method can easily be generalized to more than two strata. The designation of how many strata to use is part of the sample design and some guidelines on stratification are presented in the next section.

DESIGNING THE SAMPLE

In any survey, careful preparation of a valid sampling plan is vital. If the data that are collected do not reflect the true agricultural status, then no amount of analysis will provide a reliable crop assessment. This section deals with:

- Relationship between sample size and uncertainty in the estimate.
- Stratification and method of sampling.
- Delineation of the growing area.

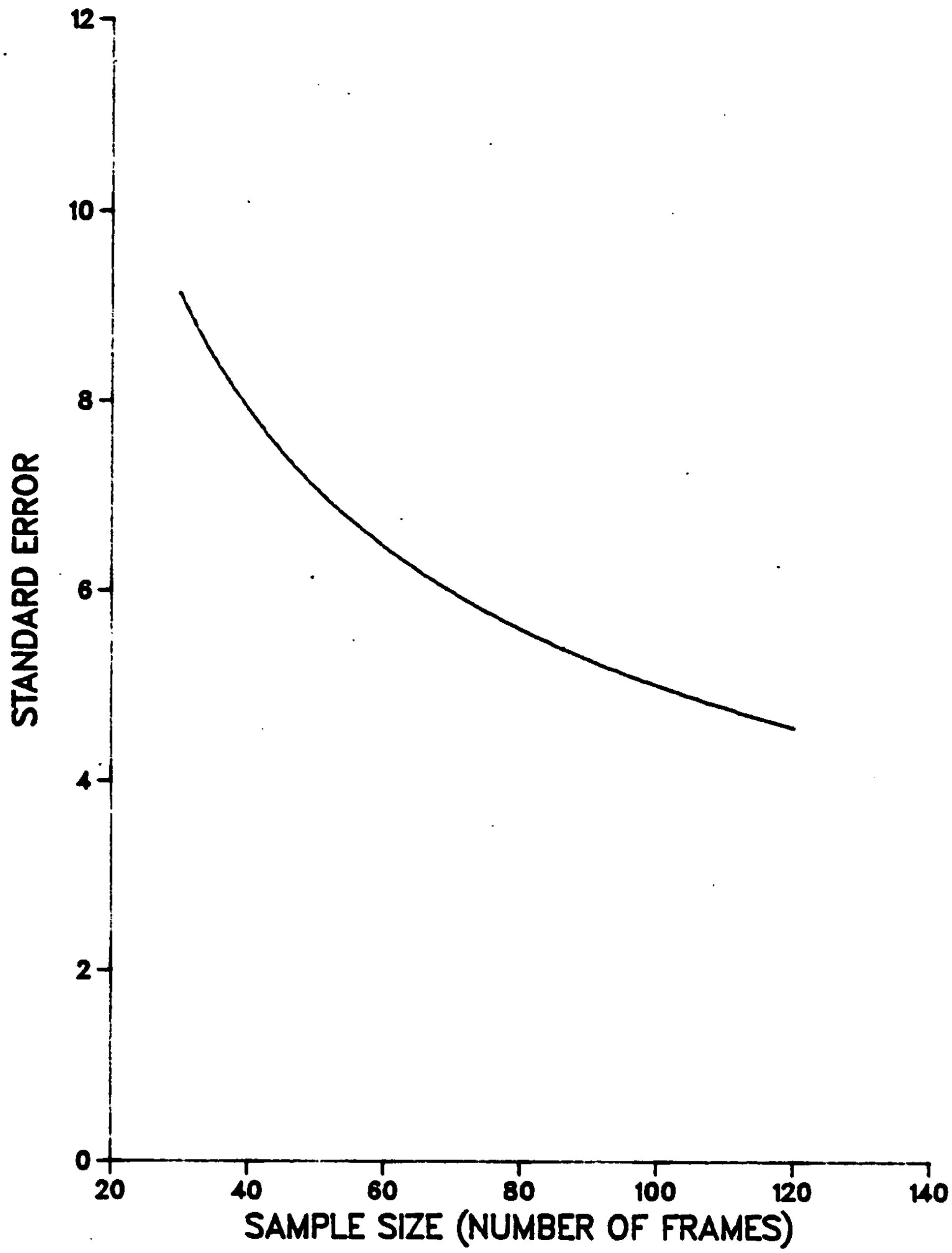
All things being equal, the more imagery that is acquired the better, since the more imagery that is exploited the smaller the uncertainty in the final esti-

25X1

mate. At some point, however, the cost of exploiting more imagery outweighs the benefit of the additional decrease in uncertainty. The uncertainty in the final estimate is a function of several factors, including sample size (figure 2). For example, doubling the sample from 50 to 100 frames would reduce the standard error, which is the statistical measure of uncertainty, from about 7.5 to 5, a gain of 50 percent. In general, the number of frames required to achieve a certain level of precision, say to estimate total area under wheat production to within plus or minus 500 hectares, is a function of the variability in the estimate of average field size and the variability in density between frames. It is often a good idea to obtain preliminary estimates of these quantities, by exploiting a few frames of imagery, from which the statistician can calculate the number of frames required to achieve the desired precision in the final estimate.

In some cases, it may be desirable to use a stratified random sample. In stratified random sampling, the country or agricultural area of interest, is divided into nonoverlapping units called strata. Then a random sample is drawn from each strata. Stratification is appropriate when:

FIGURE 2. UNCERTAINTY IN FINAL ESTIMATE VS SAMPLE SIZE



- Estimates of cultivation are desired for certain predetermined strata, such as provinces of a country.
- Administrative convenience may dictate the use of stratification. For example, different types of imagery may be available for different strata.
- Stratification will reduce uncertainty in the final estimate of cultivation if the entire country is agriculturally heterogeneous, and can be divided into homogeneous strata.

When stratification is used, the number of strata and the number of frames that should be selected per stratum must be determined. In practice, the number of strata is usually chosen via personal judgment and is in the range of two to six. The determination of how many observations to take per stratum depends on the variance in density within that stratum. For a desired precision, the optimal sample size per stratum can be calculated from a small preliminary sample. The general guidelines for determining the sample size in a given stratum are to take more observations if the stratum is larger, the stratum is more variable internally, or sampling less costly in that stratum.

Although it is important to sample enough frames of imagery, the method of sampling is equally important. Incorrect selection of targets can seriously bias the final estimate. Three major sources of bias in agricultural crop estimation are non-random selection of targets, incorrect delineation of the potential growing area, and stratification of the region based on data contained in the sample.

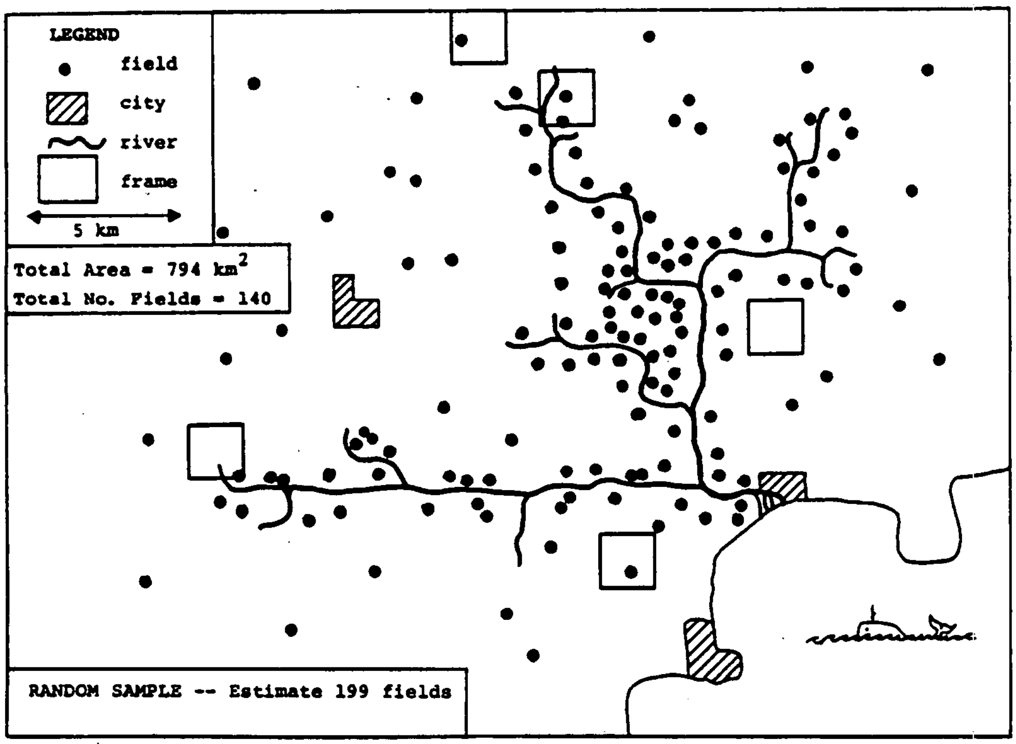
When the targets are selected randomly over the growing area, the imagery is likely to reflect the overall growing conditions. If targets are selected in a non-random fashion, such as targeting only the areas reported to have the heaviest cultivation, then a substantial bias arises. The true number of fields on the map (figure 3) is 140 and the estimate based on 5 randomly allocated frames is 199 fields. Using a biased sampling procedure and the same number of frames, the resulting estimate is 1,032 fields. Clearly, targeting only the densest region has biased the analysis beyond repair.

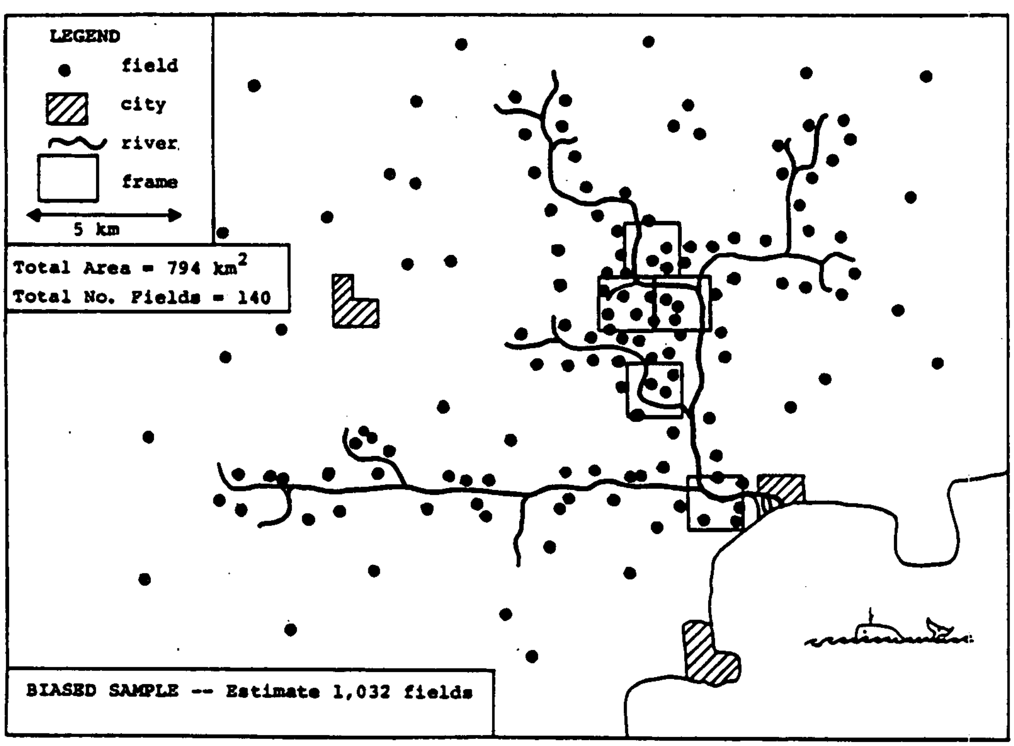
Incorrect delineation of the growing area can cause a similar problem. In figure 4, the dashed line represents growing area, as designated at the time the targets were selected. When no fields were observed on two frames, one

SECRET

25X1

FIGURE 3. RANDOM VS BIASED SURVEY

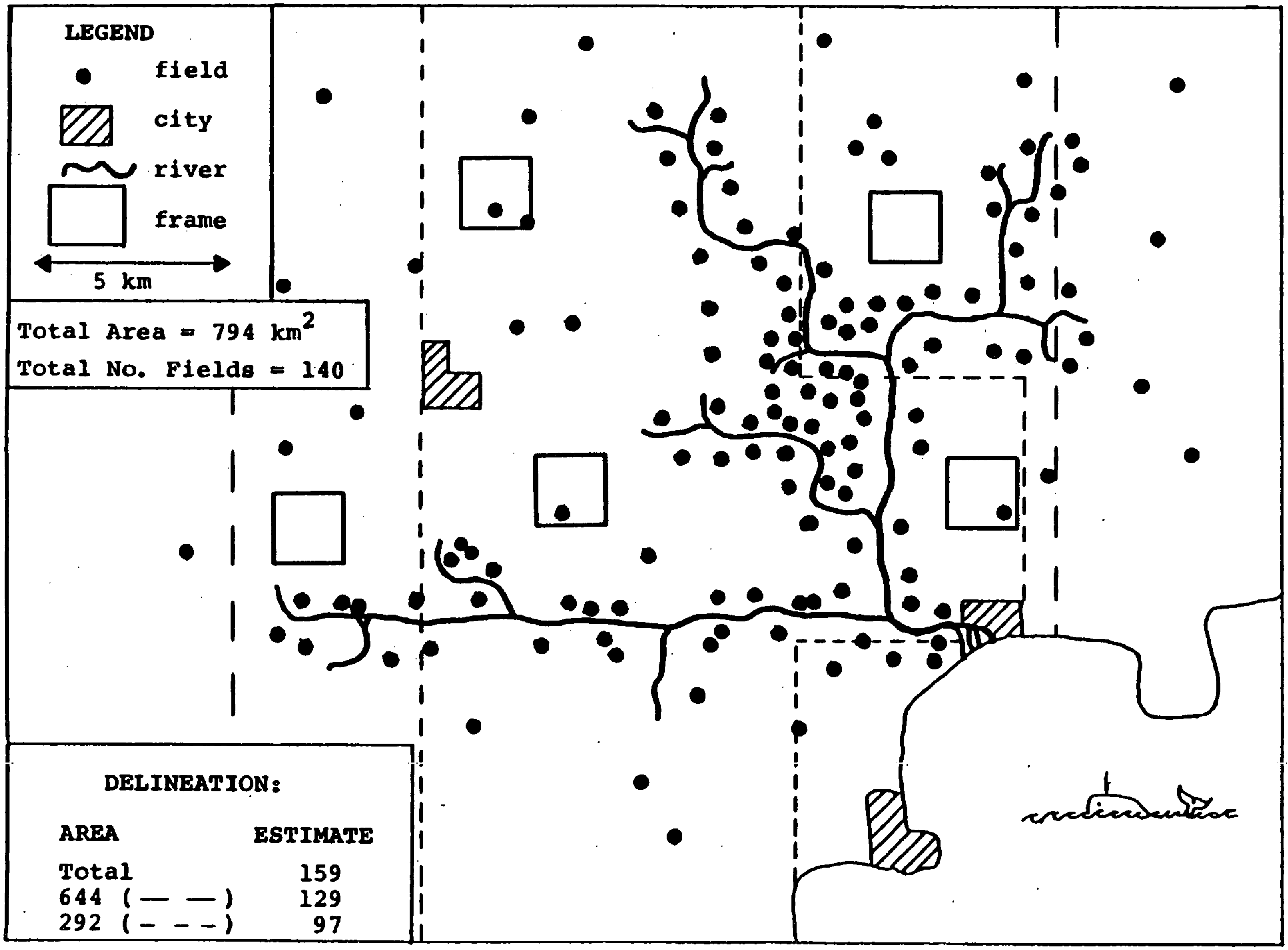




SECRET

25X1

FIGURE 4. DELINEATION OF POTENTIAL GROWING AREA



Sanitized Copy Approved for Release 2011/02/22 : CIA-RDP86T01017R000707650001-5

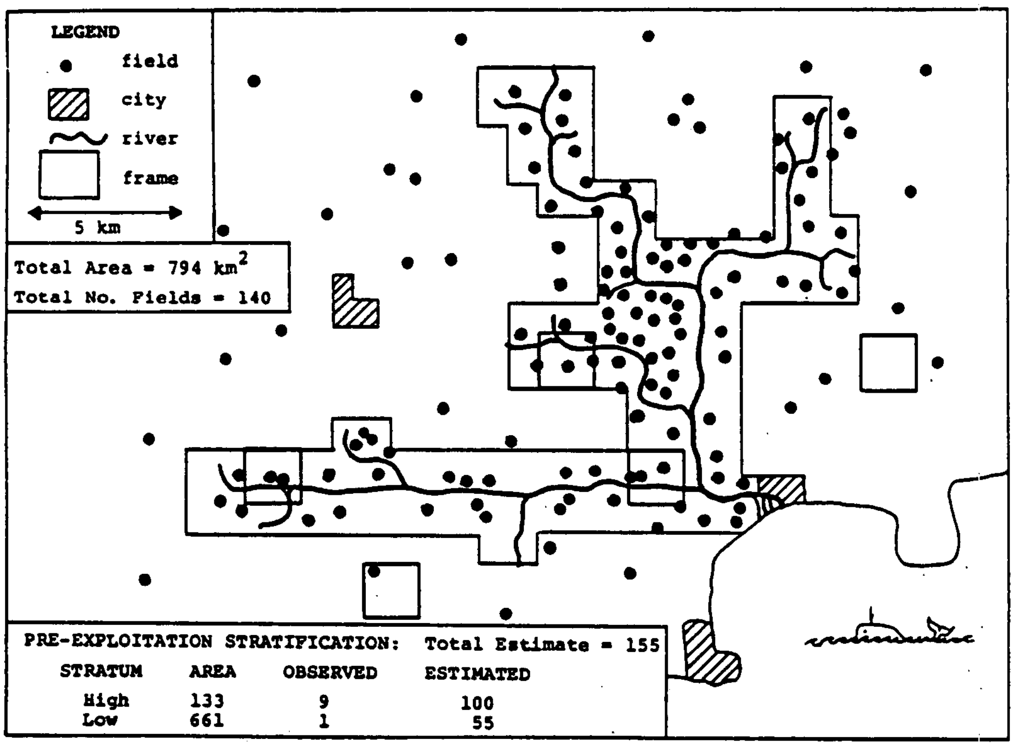
might be tempted to reduce the potential growing area as indicated by the dotted line. This would reduce the estimate from 129 to 97 fields, when the true value is 140. Thus, errors in delineation can introduce a serious bias in the final estimate.

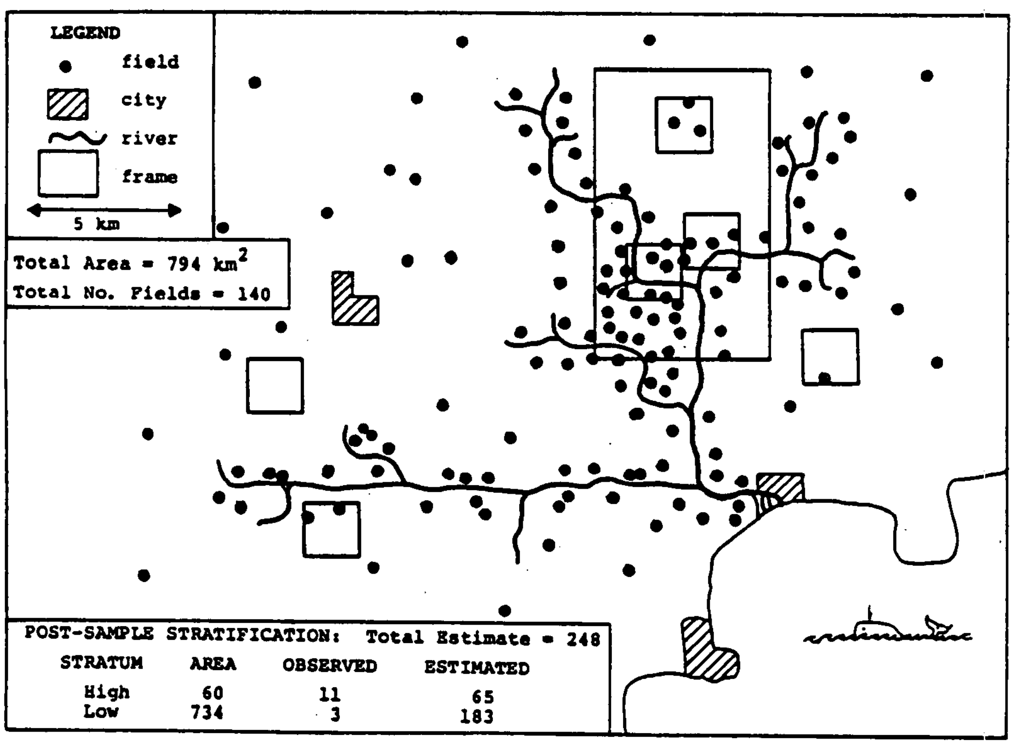
Finally, improper stratification can lead to estimation bias. It is not necessary to stratify the growing region, but if the decision is made to stratify, the allocation of areas to the different strata must be made without being influenced by the imagery from the sample. Figure 5a shows a stratification based on the river system and a corresponding sampling plan. In figure 5b, stratification was made after the imagery was examined and the high density stratum was based on the densities observed on imagery. In the first instance the estimate was 155 fields, while in the second case the estimate was 248. Since the true value is 140, the stratification which was not based on the observed densities produced a more accurate estimate of the total number of fields.

While these examples are hypothetical, they illustrate some of the problems that can arise in sampling and illustrate the need for a good sample design. If a

25X1

FIGURE 5. STRATIFICATION





proper design is used, the estimate of cultivation can be quite good.

SPECIFIC AREAS OF INTEREST

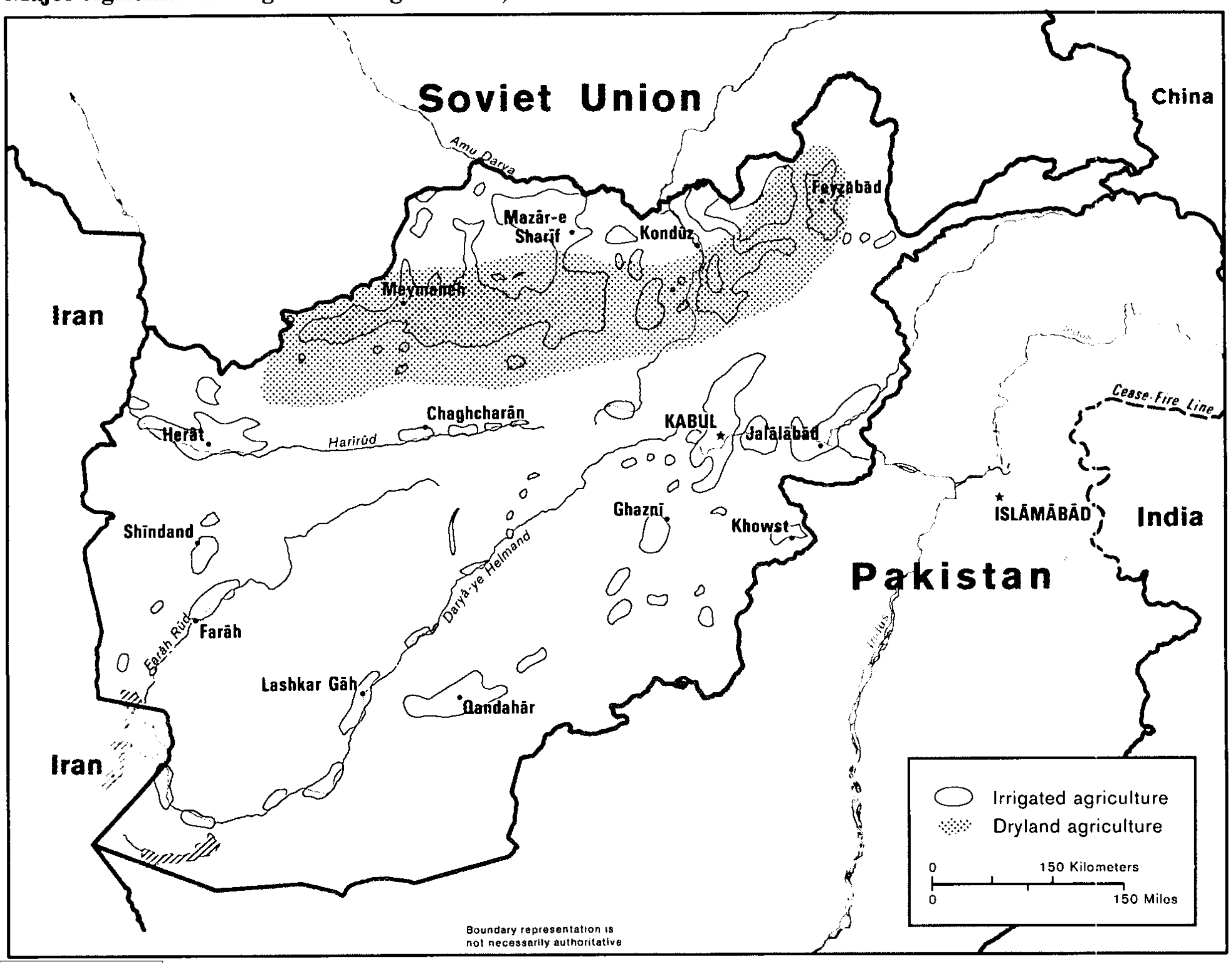
AFGHANISTAN

Assessment of grain production in Afghanistan has been of particular interest since the Soviet invasion in 1979. Grains account for 90 percent of the cropped area, with wheat being predominant. Other grain crops include, rice, barley, and corn. Most of Afghanistan consists of mountains, desert, and forest, and is unsuitable for agriculture. The productive soils account for only about 15 percent of the land and cultivation is limited to about half the arable area (about 4 million hectares). During recent years, there have been no reports of significant flooding or droughts which would cause extensive crop damage. Nevertheless, the outflow of the population from the rural areas has caused strains on the resources necessary for agricultural production.

Extensive satellite imagery is available for most parts of Afghanistan (figure 6). A comparison of the density of straw shocks indicates the relative changes in yields. Although travel to the growing regions is not

25X1

Figure 6
Major Agricultural Regions in Afghanistan, 1986



708132 (A01776) 7-86

25X1

25X1

possible, ground truth photography can be acquired in adjacent parts of Pakistan. In addition, data published prior to the Soviet invasion provides a baseline for crop comparisons.

ETHIOPIA

Farming practices in northern Ethiopia are primitive and, even with the best weather conditions, yields are not high. The drought since 1979 poses difficult problems since official Ethiopian government statistics show no change in crop production during this period when the harvest was clearly declining sharply. The major crops grown in this area are teff,[3] sorghum, and barley. Planting of crops begins in April or May and imagery suitable for crop analysis must be obtained between June and September.

In northern Ethiopia, the crops can be classified as good, fair, or poor, on the basis of vigor analysis. Tonaldifferences in the imagery reveal the crop vigor, which corresponds to different yields. By estimating the

[3]Teff is a cereal unique to Ethiopia.

hectareage separately for each category and combining these with the appropriate yields, fairly precise production estimates are possible.

CUBA

Citrus production in Cuba has expanded in recent years, as new orchards have begun to bear fruit. The analysis of this citrus crop differs from the grain assessments described above, in several respects. Since citrus orchards last for years, the concept of a crop calendar and the need to image only during a critical window are not applicable. Also, the potential growing area is relatively small, since Cuba is not a large country. Consequently, it is possible to image the entire growing region and conduct a census. This approach, while time-consuming, eliminates the uncertainties due to sampling. Problems of crop identification remain, though, and it simply is not possible to distinguish among oranges, lemons, limes, and grapefruit. The final crop assessment is, therefore, for citrus in general.

[redacted] 25X1

SOVIET UNION

Wheat production in the Soviet Union, unlike grain production in many less developed countries, has been studied extensively. Project UPSTREET, initiated in 1973, is a multidisciplinary study of grain production in the Soviet Union (reference 11). Extensive research has produced an elaborate crop model, which predicts grain yields on a regional basis using data from imagery [redacted] 25X1 The crop calendar and farming 25X1 practices are thoroughly documented for the Soviet Union. Because sufficient resources have been dedicated to the problem, Soviet grain estimates embody a degree of precision that will not be realized in most other crop assessments.

COLLECTION GAPS AND OPPORTUNITIES

One of the most difficult aspects of the crop assessment process is accurate delineation of the potential growing area. Since the sample is designed in accord with a specified potential growing area, errors in delineating this region will degrade the entire process. Use of LANDSAT imagery can alleviate some of the problem.

25X1

By collecting LANDSAT imagery over the regions of interest, it is possible to distinguish agricultural areas from other parts of the country. Thus, LANDSAT data could be used to delineate the agricultural region directly. Treating these regions as the potential growing area would facilitate efficient sample design and improve the accuracy of the final crop estimate.

Finally, LANDSAT imagery offers some indications of crop yields. Application of LANDSAT imagery to vigor analysis in the Soviet Union is already standard practice. However, variations in the tonal qualities of LANDSAT imagery may be a good indicator of moisture stress in other areas, even if the actual fields are smaller than the resolution of the LANDSAT image. For high-priority areas in the developing world, investigation of the use of LANDSAT for crop assessment is clearly warranted.

APPENDIX A

CONFIDENCE INTERVALS

The confidence interval gives an indication of the precision of the estimation process. The true amount of cultivation is a fixed but unknown number and the value obtained is only an estimate, based on a survey. The error in the measurement process will affect the final value. The 95 percent confidence interval is an assessment of this error in the estimation process. If the entire survey and estimation process were repeated many times, a collection of 95 percent confidence intervals would be produced. Of these intervals, 95 percent would cover the true but unknown size of the cultivated area being estimated.

Calculating the estimate and confidence interval is a complicated process, although conceptually it is straightforward. The growing region can be viewed as an area composed of a number of hypothetical "frames". When portions of the region are imaged, this is equivalent to sampling some of these frames. The variance of the estimate for the total number of fields reflects the variation in the density of fields and the variation due to sampling.

To demonstrate the calculations, let:

A_g = Total growing area

A_i = Total area imaged

n = Number of frames imaged

S = Average size of frames imaged (= A_i / n)

x_i = Number of fields observed on frame i

a_i = Area imaged on frame i

X = Number of fields observed on imagery

a = Average field size, based on mensuration of the imagery

V_a = Variance of a, also based on mensuration of the imagery

V_d = Variance (between frames) of the density of fields[3]

Then the number of "frames" in the total growing area is $N = A_g n/A_i$. The initial estimate of the number of fields in the growing area is F, where

$$F = A_g (X / A_i)$$

and the variance of F, call it V_F, is given by

[3]If the area imaged on each frame has been mensurated, then direct calculation of V_d is possible. Otherwise, it must be approximated from the variance in the number of fields across frames and the average frame size--a procedure which will inflate the final variance estimate.

25X1

$$V_F = V_d \frac{A_g^2}{A_i^2} \sum_{i=1}^{n} a_i^2 \left(1 - \frac{n}{N}\right).$$

The ground truth information permits estimation of the detection probability and F is adjusted accordingly. Let:

x_g = The number of fields observed in the ground truth.

x_i = The number of fields observed in the same area on imagery.

Then

ρ = true detection probability

is estimated by

$p = x_i / x_g$

and the variance of p, call it V_p, is given by

$V_\rho = \rho(1-\rho)/ x_g$.

Sanitized Copy Approved for Release 2011/02/22 : CIA-RDP86T01017R000707650001-5

25X1

The estimated number of fields, adjusted for detection errors, is F_a where

$$F_a = F / \rho$$

and the variance of F_a is

$$V_{F_a} = \frac{1}{p^4} \{ V_p V_F + F^2 V_p + p^2 V_F \} .$$

The above formula is an approximation. Since $F_a = F / \rho$, it is necessary to calculate the variance of $1/p$. A Taylor series expansion of $1/\rho$ about the point p yields

$$1/\rho = p/\rho^2 + O(\rho^2)$$

Ignoring terms of order ρ^2 and higher leads to the variance calculation shown above.

To complete the calculations, the estimated area cultivated is

$$A_c = a F_a$$

and the variance of the estimate is

25X1

$$VA_c = V_a\ VF_a + a^2\ VF_a + F_a^2\ V_a$$

To derive the confidence interval, use the appropriate multiple of the standard deviation of the estimate. The standard deviation of A_c is the square root of VA_c. So, the approximate 95 percent confidence interval is

$$A_c \pm 2\ (VA_c)^{1/2} \quad .$$

25X1

APPENDIX B

DATA REQUIREMENTS CHECKLIST

Sanitized Copy Approved for Release 2011/02/22 : CIA-RDP86T01017R000707650001-5

[redacted] 25X1

This section presents a checklist of the primary statistical issues involved in the preparation of an estimate. As such, it does not present any new methodology but should assist the analyst in planning for a new crop estimate.

SAMPLING PLAN

Development of a sampling plan involved three steps: delineation of the growing area, the decision of whether or not to stratify, and selection of the targets. The delineation of the growing area should rely on [redacted] 25X1
[redacted] analysis of the regions which are suitable for 25X1
cultivation of a crop and, if possible, imagery from previous years. Once the potential growing area has been delineated, the decision must be made whether or not to stratify. It is important that the decision to stratify be made without being influenced by the imagery at hand. Finally, the targets should be chosen randomly within the potential growing region or within the strata.

GROUND TRUTH

Ground truth photography is not always available, due to time and cost limitations, in addition to the fact that

the region of interest may be a denied area. Nevertheless, ground truth should be obtained whenever possible. To develop "keys" for examining the overhead imagery, it is essential to acquire ground truth photography and overhead imagery of the same location at approximately the same time. These two sets of imagery are also used in the calibration experiment described below. Finally, the ground truth trip offers the analysts an opportunity to discuss the crop conditions with field personnel [redacted] 25X1
25X1

IMAGERY EXPLOITATION

Exploitation of the imagery is the most time consuming step in the estimation process. Ideally, the data derived from imagery should include the number of fields and the area imaged on a frame-by-frame basis. The proportion of the area obscured by clouds must also be recorded. Average field size is determined by mensurating a sample of the fields observed on the imagery.

CALIBRATION EXPERIMENT

The calibration experiment provides an estimate of the detection and misclassification probabilities.

Ideally, these quantities are estimated by exploiting selected imagery and comparing the results to ground truth data. In the absence of ground truth, a calibration experiment can be performed using two imagery analysts and a technique known as capture-recapture modeling.

YIELDS

To estimate the total area under cultivation requires only the information listed above, but to assess production one must have information on yields. Generally, this information [redacted] 25X1 is the weakest aspect of the analysis. 25X1

Next 1 Page(s) In Document Denied

Sanitized Copy Approved for Release 2011/02/22 : CIA-RDP86T01017R000707650001-5

Sanitized Copy Approved for Release 2011/02/22 : CIA-RDP86T01017R000707650001-5